

# Measurement II: An Overview of “Recent” Advances

William F. Chaplin, Ph.D

Dept of Psychology

St. John's University

New York, New York

# Generalizability Theory

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 16, 137-163.

Reliability and Validity are typically viewed as discrete concepts in classical psychometrics.

G-Theory blends these concepts together by recognizing that the fidelity with which we can *generalize* a particular measure to a broader *universe* of measures is our goal.

How well can we generalize a diagnosis of hypertension made by one MD to other MD's?

How well can we generalize a depression score based on one measure of depression to other measures of depression?

**These questions blend issues of “reliability” (inter-judge agreement, alternate forms) with validity (hypertension, depression).**

**G-theory is complex and is based analytically on repeated measures ANOVA designs. However, most simply it can be related to the general concern of all research: “How confidently can I generalize my results beyond the specific experiment?”**

A readable introduction can be found in

Shavelson, R. J., & Webb, N. M. (1991).

*Generalizability theory: A primer.* Newbury Park, CA: Sage.

What you may not know about coefficient alpha.

1. It is NOT a reliability coefficient in the classic sense. It is a generalizability coefficient. (Influenced by both measurement error AND item content.)

2. Alpha is NOT a measure of a scales unidimensionality, rather the legitimacy of alpha as a psychometric measure assumes unidimensionality

3. Alpha depends upon the number of items a scale contains as well as their consistency.

4. A high alpha can “paradoxically” reduce a scale’s validity. Called the “attenuation paradox.”

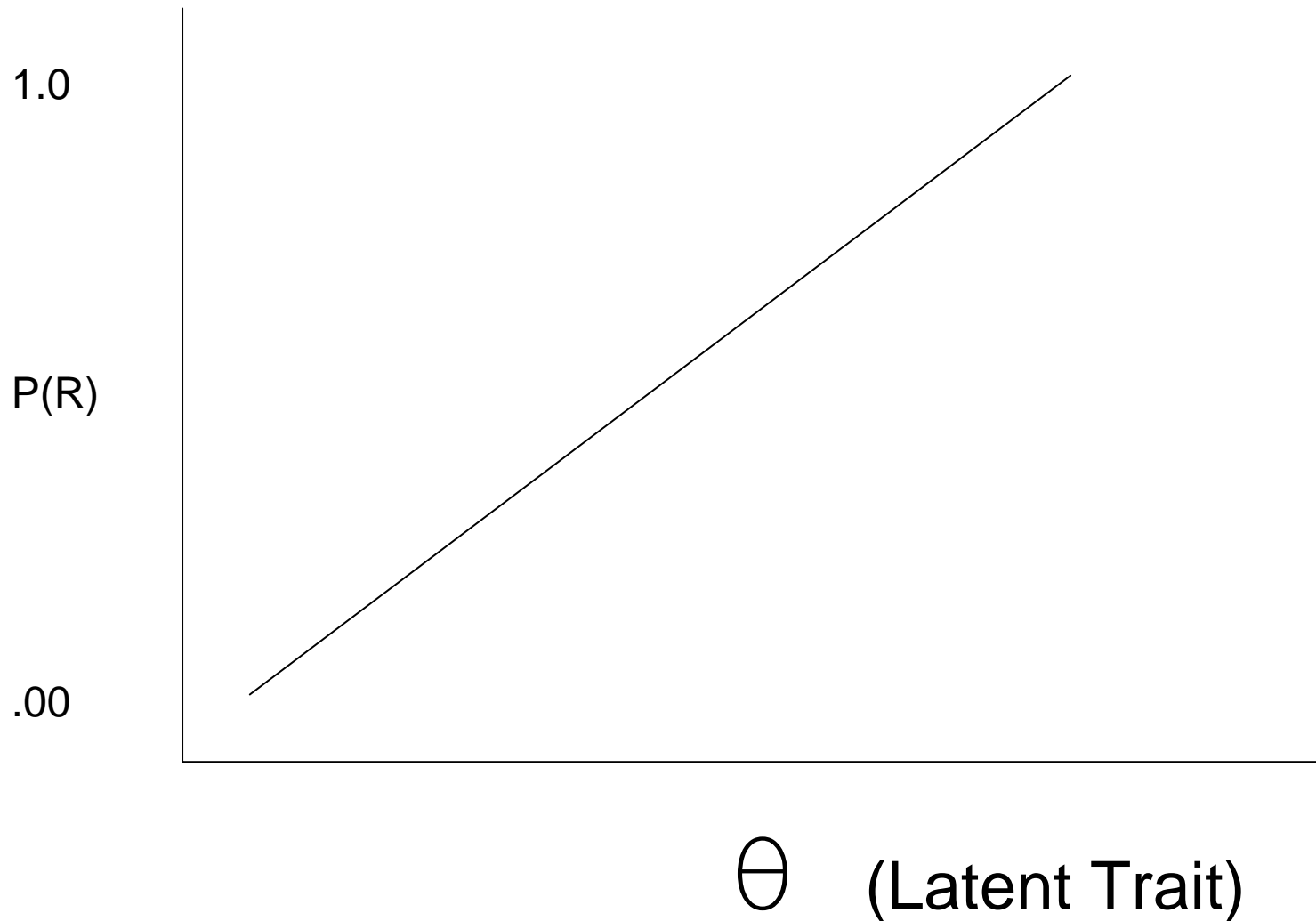
(See assigned Chapter by John and Benet-Martinez)

# Item Response Theory (IRT)

All measurement ultimately is based on our theories of how a person's observed response to an item is related to the underlying characteristic that we are trying to measure.

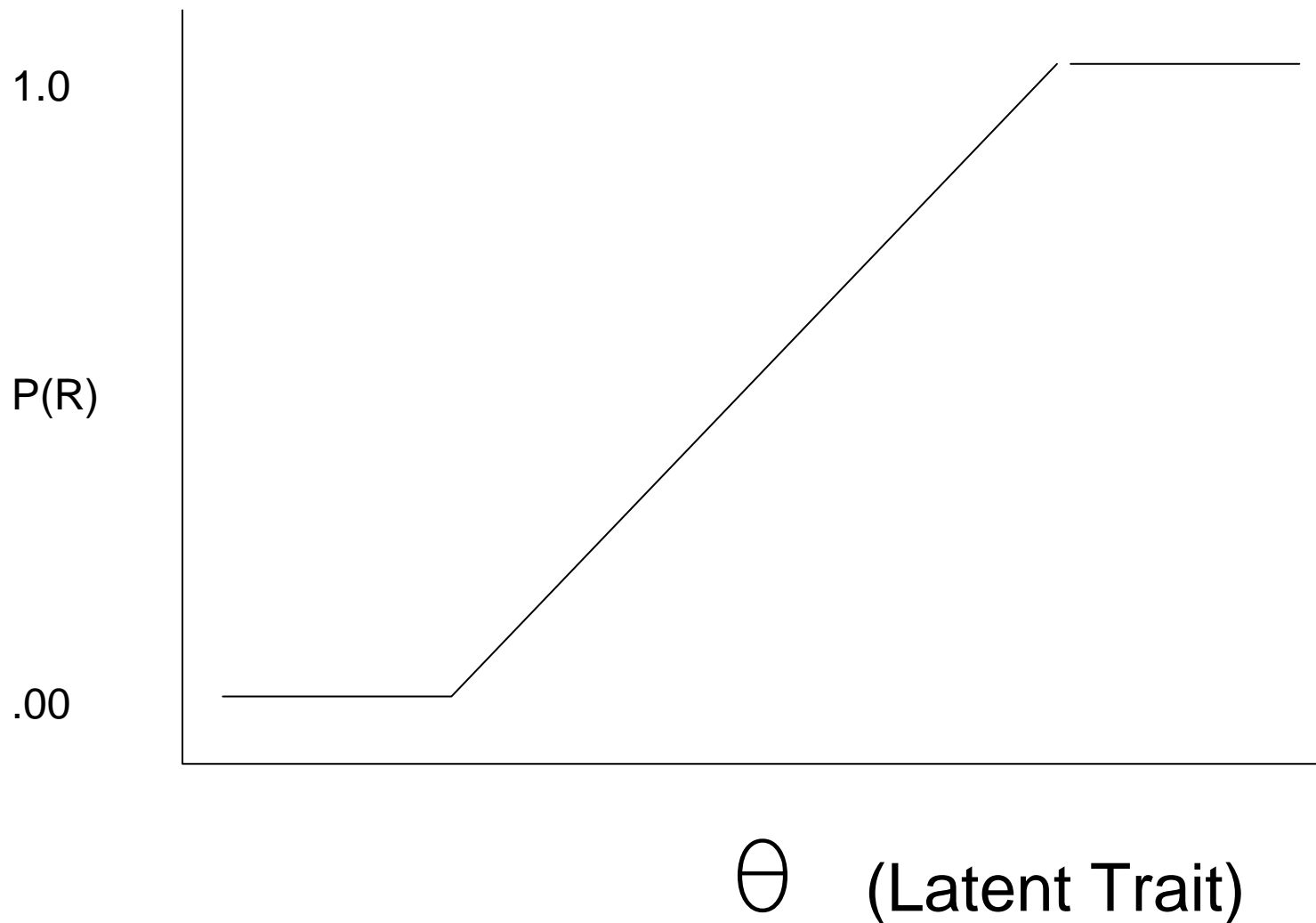
In everyday measurement these theories of item response are often not specified. IRT seeks to make explicit and specific what has historically been implicit and vague.

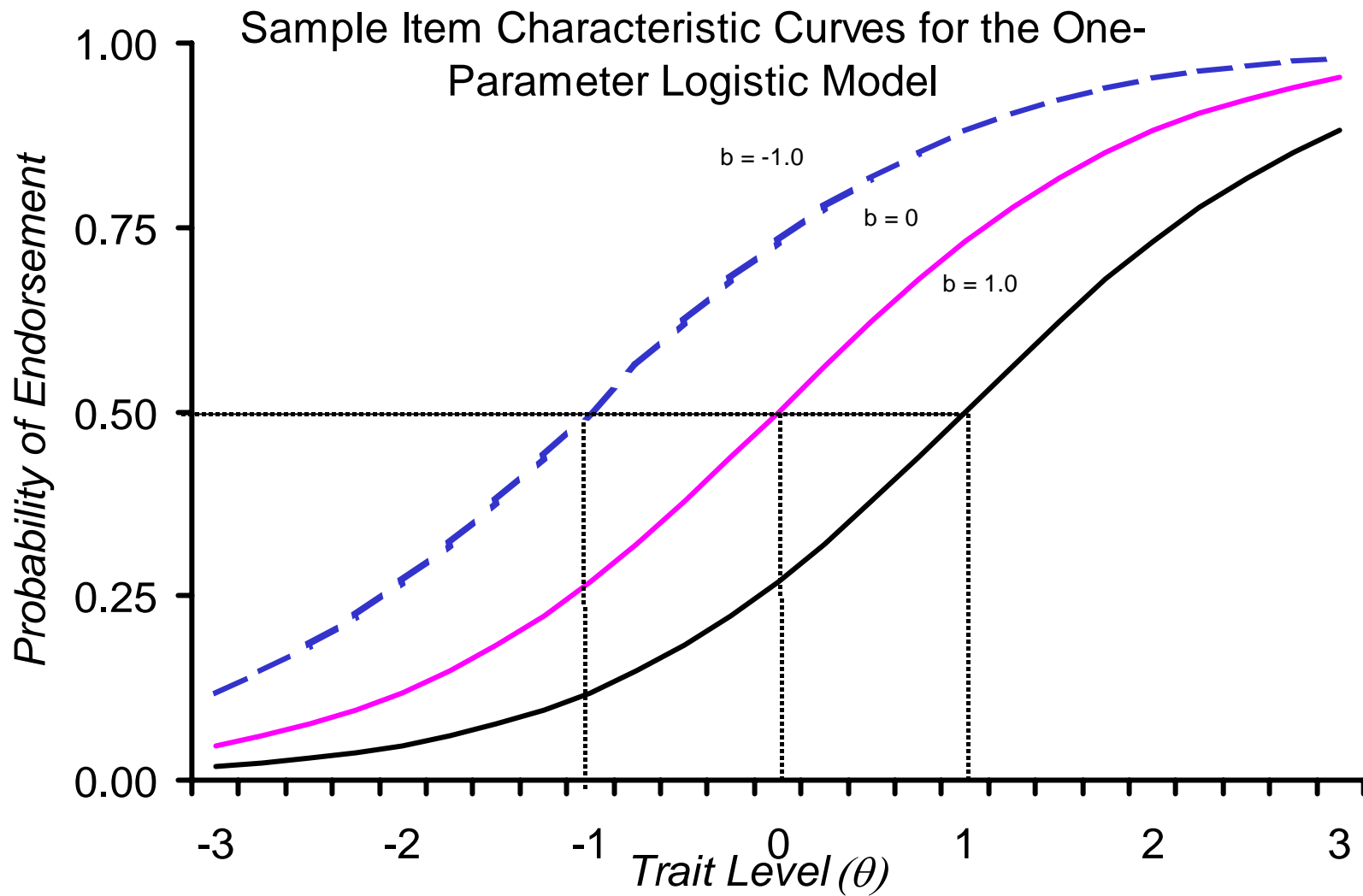
# Item Characteristic Curve (ICC)



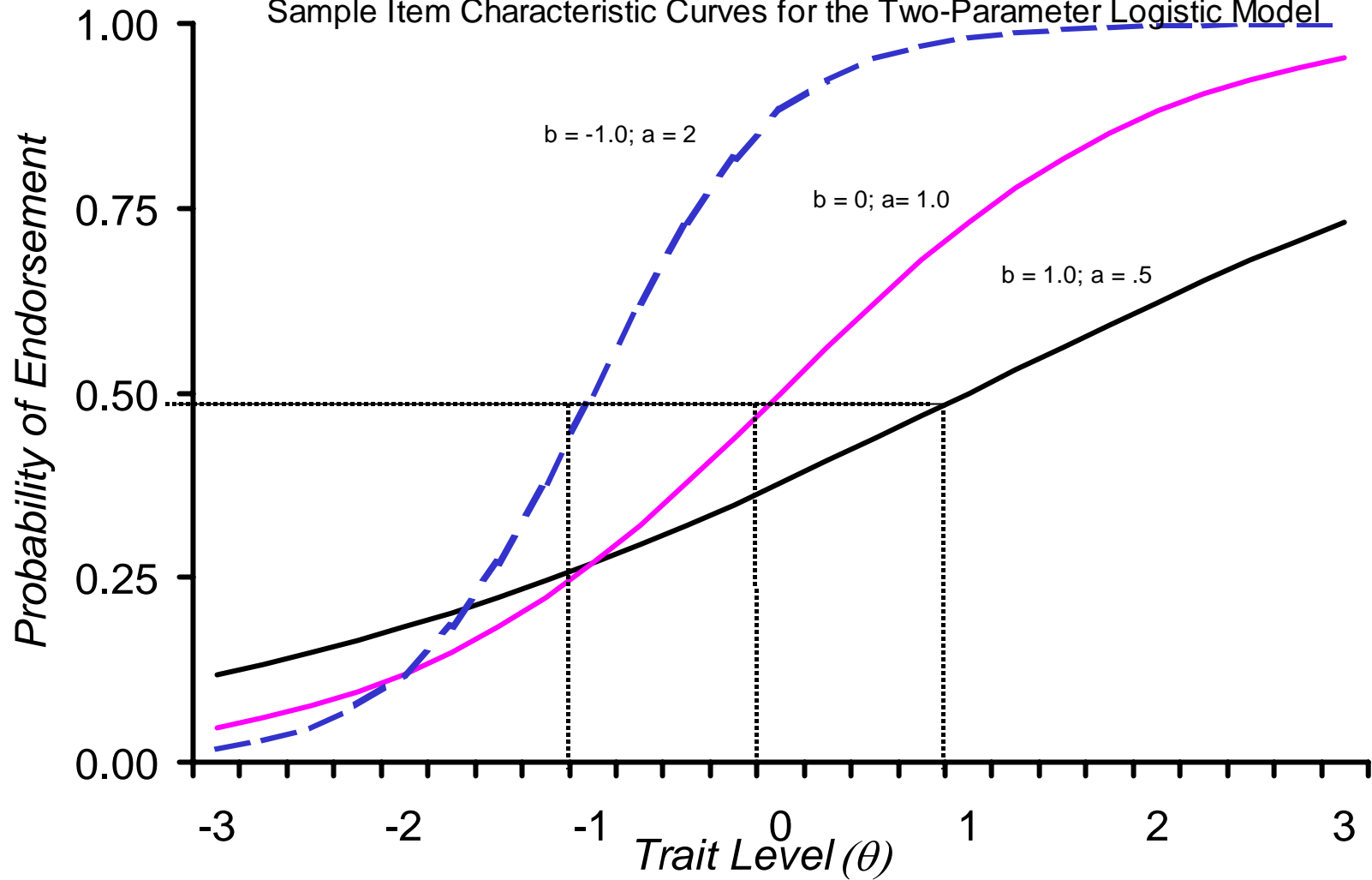


# Item Characteristic Curve (ICC)

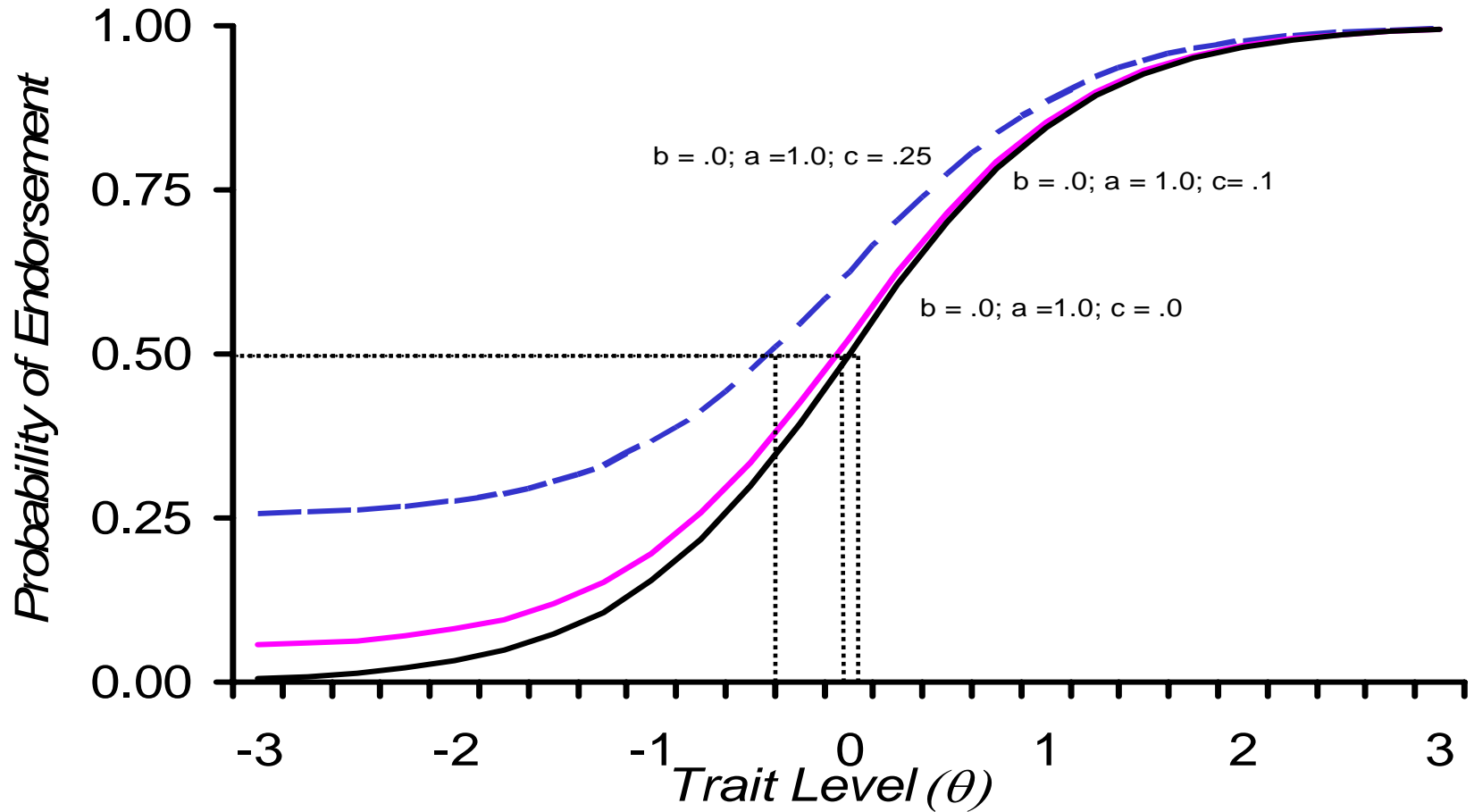




Sample Item Characteristic Curves for the Two-Parameter Logistic Model



Sample Item Characteristic Curves for the Three-Parameter Logistic Model



# 1. Three-Parameter Logistic Model

$$P(X_{is} = 1 \mid \theta_s, b_i, a_i, c_i) = c_i + (1 - c_i) \frac{\exp[a_i(\theta_s - b_i)]}{1 + \exp[a_i(\theta_s - b_i)]}$$

$X_{is}$  = response of person  $s$  to item  $i$  (0 or 1)

$\theta_s$  = trait level for person  $s$

$b_i$  = difficulty or threshold of item  $i$

$a_i$  = discrimination of item  $i$

$c_i$  = the lower asymptote of item  $i$

$\exp$  = the natural log base (2.718)

Estimation of parameters proceeds by iterative maximum likelihood procedures.

IRT provides a very strong model that can lead to precise estimates of the person's characteristic and to the item characteristics.

Used to develop measures as well as scale people.

Readable Introduction:

Hambleton, R. K., Swaminathan, H., & Rogers, L. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

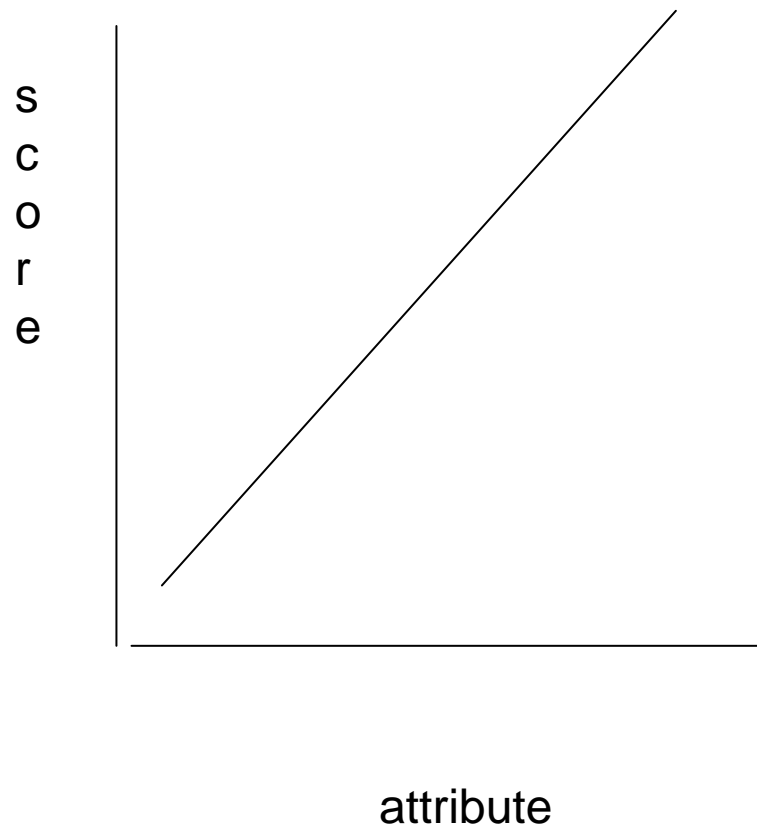
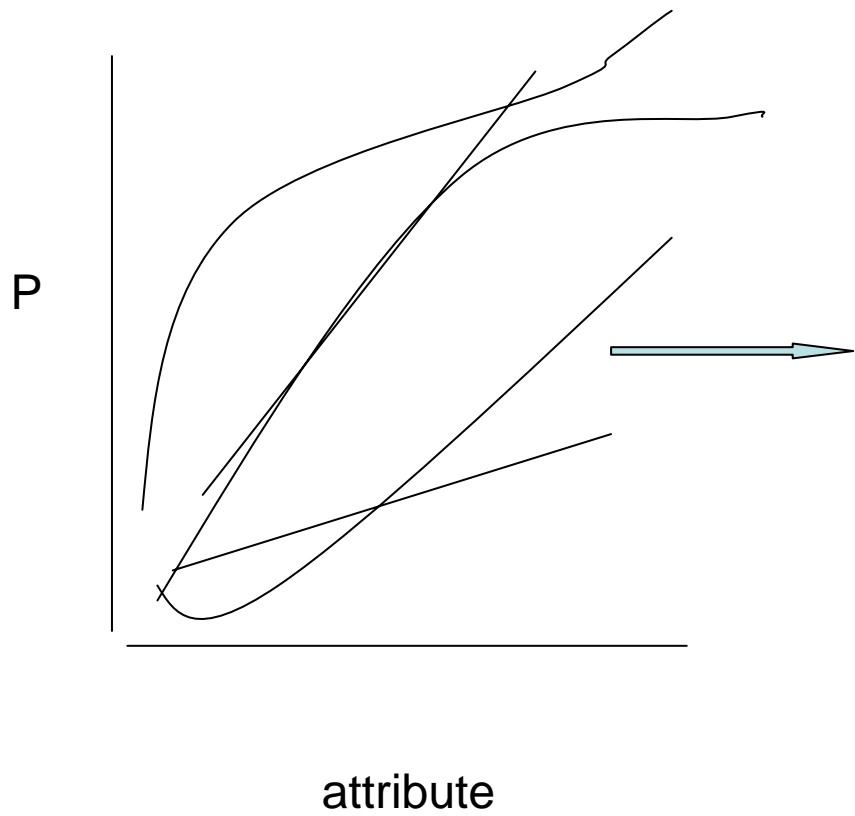
Most common scoring of a scale is to sum (or average) the items.

This is based on a powerful, but vague IRT called the “General Linear Scaling Model”

Assumptions:

Items are unidimensional

ICC's are monotonic





# Confirmatory Factor Analysis (CFA)

A form of structural equation modeling (SEM) often called the “measurement model” component of SEM.

CFA allows you to test that your items have particular structural properties such as unidimensionality

$$\begin{aligned}
 & \left[ \begin{array}{c} \text{r}_{ij} \end{array} \right] - \left[ \begin{array}{c} \text{Est r}_{ij} \end{array} \right] \\
 & = \left[ \begin{array}{c} \text{Res r}_{ij} \end{array} \right]
 \end{aligned}$$

Researchers who apply CFA techniques “do not seem adequately sensitive to the fundamental reality that there is no true model....and that the best one can hope for is to identify a parsimonious, substantively meaningful model that fits observed data adequately well.” (p. 213)

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, 51, 201-226.

## Example—MHLCs

MHLCs developed by Wallston et.al. to assess three dimensions of health locus of control beliefs Internal (I), and two External Dimensions, Powerful Others (PO) and Chance (C)

In 1996 a revised MHLCs was developed that added a third External Dimension (God (G))

We used CFA to evaluate the structure of the MHLCS to see if its structure was consistent with the hypothesized dimensions

The revised MHLCS has 24 items; 6 items on each of the four subscales

### Correlations Among the Four Health Locus of Control Subscales

Subscale		1	2	3	4
1.	Internal	---	.10	.14	.14
2.	Chance		---	.54	.50
3.	God			---	.43
4.	Powerful Others				---

-----  
Note. N = 371. Correlations larger than .11 are significant at the .05 level, two-tailed test.

Summary of the Goodness of Fit Indices for Different Structural Models of the Health Locus of Control Items

Standardized Model	Degrees of Freedom	Chi- Square	Chi-Square	S-B Scaled CFI	Robust CFI	RMSR
Independence	276	2679.7	-----	----	----	----
1 Factor	252	963.4	763.4	.704	.722	.089
4 Independent Factors	252	763.3	625.3	.787	.797	.156
4 Correlated Factors	246	490.4	399.6	.898	.916	.067
2 Independent Factors (Internal vs External)	252	764.3	610.8	.787	.805	.084
2 Correlated Factors	251	762.0	609.6	.787	.805	.082
4 Factors with three External Factors Correlated	249	492.8	401.2	.899	.917	.069

---

Note. N = 371. S-B = Sattora-Bentler; CFI = Comparative Fit Index; RMSR = Root Mean Squared Residual

Standardized Estimates of the Path Coefficients for the Four Factor Model with the  
"External" Factors Correlated for the HLCS

Item	Internal	Chance	Factor God	Powerful Others	Error
I1	.36				.93
I2	.35				.94
I3	.14*				.99
I4	.58				.81
I5	.58				.81
I6	.73				.68
C 1		.46			.89
C 2		.41			.91
C 3		.53			.85
C 4		.55			.84
C 5		.47			.88
C 6		.60			.80
G 1			.73		.68
G 2			.49		.87
G 3			.78		.63
G 4			.78		.62
G 5			.84		.54
G 6			.83		.56
P 1				.47	.89
P 2				.50	.87
P 3				.10*	.99
P 4				.61	.79
P 5				.60	.80
P 6				.72	.70

---

Correlations Among the Factors				
	Chance	God	Powerful	Others
Chance	----			
God	.71	----		
Powerful Others	.77	.56	----	

Note. N = 371. I = Internal, C = Chance, G = God, P = Powerful Others. Items are grouped by factor. The number of the item corresponds to the order it appears on the MHLCS. All paths are



Coefficient Alpha and Range of Item Total Correlations  
for the MHLCS Scales

<u>Scale</u>	<u>Coefficient Alpha</u>	<u>Range of Corrected Item-Total Correlations</u>
Internal	.60	.14 - .45
Chance	.68	.35 - .47
God	.88	.47 - .77
Powerful Others	.65	.10 - .49